ELSEVIER

Discrete Mathematics 147 (1995) 211–234

DISCRETE MATHEMATICS

Bent functions and random boolean formulas

Petr Savický*

Department of logic, Faculty of Philosophy, Charles University, Prague, 11638 Praha 1, Czech Republic

Received 28 February 1991; 16 March 1994

Abstract

Let α be a nonlinear boolean connective with equal number of zero and unit entries in its table. We study an iterative process of combining randomly chosen boolean functions from some starting set G via the connective α. We suppose that all functions in G are defined on the same finite nonempty domain M. We are interested in the situations, when the process converges to the uniform distribution on $\{0,1\}^M$. We classify the boolean connectives α according to the asymptotic rate of this convergence.

Although the probability of any function in our process converges to $(\frac{1}{2})^{|M|}$, there are some differences in the terms of lower order of magnitude. If M is the boolean cube of an even dimension n and G is the set of all linear boolean functions of n variables and the connective α belongs to the class of the lowest possible rate of convergence in the above-mentioned classification, we can express the main nonconstant term of the asymptotic expansion of the probability of occurring of a single function f through the Fourier transform of f. Using this, we prove that the bent functions achieve asymptotically the minimal probability of occurring among all boolean functions. At the same time, the linear functions achieve asymptotically the maximal probability.

1. Introduction

Let $k \geqslant 2$ be a natural number and $\alpha: \{0,1\}^k \to \{0,1\}$ be a boolean connective. Let M be a nonempty finite set and let G be a nonempty set of functions $M \to \{0,1\}$. By combining functions randomly chosen from G via the connective α, one may obtain a probabilistic distribution on a larger set of functions $M \to \{0,1\}$ than G. This step may be repeated iteratively. We are interested in the situations when this process converges to the uniform distribution on all functions $M \to \{0,1\}$.

In [7] this process is studied particularly for $M = \{0,1\}^n$, the boolean cube, and $G = \{0, 1, x_1, \ldots, x_n, \neg x_1, \ldots, \neg x_n\}$ for a natural number n. The main result of the

* E-mail: petr.savicky@cuni.cz.

SSDI 0012-365X(94)00166-9

cited paper is that for this particular G the process converges to the uniform distribution if and only if α is balanced (Definition 2.2 below) and nonlinear. It appears however that the proof of the 'if' part of this result applies immediately to a more general choice of G.

In the present paper we study the process for this more general choice of G and for balanced nonlinear connectives. These conditions imply that the process converges to the uniform distribution on the set of all functions $M \to \{0,1\}$. Our goal is to prove more detailed results on the asymptotic behaviour of the process than the study of its limit. Several statements from [7] are used without giving the proof.

We introduce a notion of a degree of a connective. It is the minimum number of variables of the connective that must be set in order to obtain a nonbalanced subfunction. Since we are studying balanced connectives, the degree of our connectives is always at least 1. We prove that the asymptotic rate of the convergence of the process described above to the uniform distribution depends substantially on the degree of α. The higher is the degree, the faster is the convergence. The precise formulation may be found in Theorems 4.7 and 4.8. Some combinatiorial properties of the degree may be found in Section 3.

In our situation the difference between the distribution in the ith step of the process and the uniform distribution tends to zero if i tends to infinity. We study in Section 5 the asymptotic behaviour of this difference if α is nonlinear and of degree one. In this case the difference may be approximated up to the terms of lower order of magnitude using the behaviour of the restrictions of the functions in G only to subsets of M of a small fixed size. The contributions of individual subsets are limits of iterations of some nonlinear recurrence relations. We have no closed simple formula expressing these limits. Hence, we introduce one more assumption. Namely, we suppose that G is a linear code, i.e. it is closed under addition modulo 2. Under this assumption, there is only one 'unexpressible' real constant in the resulting asymptotic formula.

In [6] Rothaus introduced a class of boolean functions which are called bent functions. These functions are defined using their discrete Fourier transform. A function is bent if and only if all coefficients of its Fourier transform have the same absolute value. Besides other results, Rothaus proved that such functions exist, that they have always an even number of variables and that they have the maximal possible Hamming distance from the set of linear boolean functions, i.e. the distance $2^{n-1} - 2^{n/2-1}$, where n is the number of variables. The last property can also be used as a definition of the bent functions. There are several infinite classes of bent functions given in [6]. More detailed information on bent functions can be obtained in [6] mentioned above and also in the book [5]. A simple characterization of all bent functions is not known.

Recently, there are some attempts to find generating and counting procedures for bent functions, see [1, 11]. There are also several symmetric bent functions. These functions and their generalization were used in [3] to separate the class of polynomial threshold functions from the class of functions computable by depth 2, unbounded

fan-in polynomial size circuits of linear threshold gates. All the symmetric bent functions are characterized in [8].

In Section 6 we apply the approximation of the distribution derived in Section 5 to the case when M is the boolean cube of an even dimension n and G is the set of all linear boolean functions of n variables. The main result of Section 6 is as follows: if α is nonlinear and of degree 1 and i is large enough then the probability that the random function obtained in the ith step of the studied process is equal to a given bent function is less than the probability of any given nonbent function. Moreover, the linear functions, the functions from the set G itself, have probability greater than any other function.

2. Preliminaries and the degree of a connective

Let g be a positive real function defined on natural numbers. We denote by $\Omega(g)$ an arbitrary function f satisfying $\liminf_{i\to\infty} f(i)/g(i) > 0$. Moreover, we denote by $\Theta(g)$ an arbitrary function f satisfying $f = O(g)$ and $f = \Omega(g)$.

For every finite set X, let $\mathscr{B}_X = \{0,1\}^X$ and let 0_X be the zero function on the set X. In particular, $\mathscr{B}_k = \{0,1\}^k$. If $u \in \mathscr{B}_X$ then $|u|$ denotes the weight of u, i.e. the number of positions of u which are equal to 1. If $u, v \in \mathscr{B}_X$ then $u \leqslant v$ means that $u(x) \leqslant v(x)$ for all $x \in X$ and $u < v$ that $u \leqslant v$ and $u \neq v$. We denote by $\oplus$ the addition modulo 2 and by $\langle u, v\rangle$ the scalar product of u and v modulo 2.

Let M be an arbitrary nonempty finite set and let $\alpha: \mathscr{B}_k \to \{0,1\}$, $k \geqslant 2$ be a nonlinear boolean connective. The elements of $\mathscr{B}_M$ will be called boolean functions, or simply, functions. Let G be an arbitrary nonempty finite subset of $\mathscr{B}_M$. The set G will be called the starting set.

We suppose for simplicity that α, M and G are fixed in the rest of this section. The symbols f, g, u, v, w possibly with indices are used to denote functions from $\mathscr{B}_M$. The symbol $\tilde{v}$ is an abbreviation for $[v_1, \ldots, v_k]$ and analogously for f instead of v.

Definition 2.1. (a) Let $\tilde{g}_0$ be a random variable with the uniform distribution on the starting set G.

(b) For every $i \geqslant 0$, let $\tilde{g}_{i+1} = \alpha(\tilde{g}_{i,1}, \ldots, \tilde{g}_{i,k})$, where $\tilde{g}_{i,j}$ are independent realizations of $\tilde{g}_i$.

(c) For every $f \in \mathscr{B}_M$, let $p_i(f) = P(\tilde{g}_i = f)$.

It is easy to see that $\tilde{g}_i$ can also be obtained in the following way: take a formula constructed from the connective α such that the inputs of the formula are all of depth i and substitute the independent realizations of $\tilde{g}_0$ for the inputs. Hence, $\tilde{g}_i$ corresponds to a random boolean formula of depth i. The reader is referred to [10] for more information on boolean formulas.

We are interested in the situations when p_i, the distribution of $\tilde{g}_i$, tends to the uniform distribution on $\mathscr{B}_M$. If $\tilde{g}_i$ tends to the uniform distribution and $a \in M$, then

both $\tilde{g}_i(a), \tilde{g}_{i+1}(a)$ take the values 0 or 1 with probability $\frac{1}{2} + o(1)$. Hence, it is an easy consequence of (b) in Definition 2.1 that α must be balanced in the following sense.

Definition 2.2. The connective α is balanced if $|\{t \in \mathscr{B}_k \mid \alpha(t) = 1\}| = 2^{k-1}$.

Remark. There is no binary balanced nonlinear connective. There are such ternary connectives, e.g. $xy \oplus z$, $T_2^3(x, y, z) = xy \vee yz \vee xz$ and $\mathrm{sel}(x, y, z)$ defined as $\mathrm{sel}(0, y, z) = y$ and $\mathrm{sel}(1, y, z) = z$. This may be verified by a straightforward computation.

For the study of p_i we use its discrete Fourier transform defined by the following relations:

$$\Delta_i(w) = \sum_{f \in \mathscr{B}_M} p_i(f)(-1)^{\langle w, f \rangle}, \tag{1}$$

$$p_i(f) = \frac{1}{2^{|M|}} \sum_{w \in \mathscr{B}_M} \Delta_i(w)(-1)^{\langle w, f \rangle}. \tag{2}$$

Since $p_i(f) \geqslant 0$ for all f and the sum of $p_i(f)$ over all f is equal to 1, we have $\Delta_i(0_M) = 1$ and $|\Delta_i(w)| \leqslant 1$ for all w and $i \geqslant 0$. Notice that if $w \neq 0_M$, then $\sum_f (-1)^{\langle w, f \rangle} = 0$. It follows from this and the formulas (1) and (2) that the distribution p_i tends to the uniform distribution on $\mathscr{B}_M$ if and only if $\Delta_i(w)$ tends to zero for all nonzero w. Moreover, the following bounds hold.

Lemma 2.3. *For all $i \geqslant 0$ we have*

$$(\tfrac{1}{2})^{|M|} \max_{w \neq 0_M} |\Delta_i(w)| \leqslant \max_{f \in \mathscr{B}_M} |p_i(f) - (\tfrac{1}{2})^{|M|}| \leqslant \max_{w \neq 0_M} |\Delta_i(w)|. \tag{3}$$

Proof. It is an easy consequence of the following two facts.

For every $w \neq 0_M$,

$$\Delta_i(w) = \sum_{f \in \mathscr{B}_M} (p_i(f) - (\tfrac{1}{2})^{|M|})(-1)^{\langle w, f \rangle},$$

and for every $f \in \mathscr{B}_M$,

$$p_i(f) - (\tfrac{1}{2})^{|M|} = \frac{1}{2^{|M|}} \sum_{w \neq 0_M} \Delta_i(w)(-1)^{\langle w, f \rangle}. \qquad \square$$

Using this, we may obtain the estimates of the difference between p_i and the uniform distribution on $\mathscr{B}_M$. In order to do it, we shall also derive a system of recurrence relations for p_i and Δ_i.

Due to the independency of $\tilde{g}_{i,j}$ in the definition of $\tilde{g}_{i+1}$ we have

$$p_i(f) = \sum_{\alpha(f_1, \ldots, f_k) = f} p_i(f_1) \cdots p_i(f_k),$$

where the summation is over all k-tuples $[f_1, \ldots, f_k]$ satisfying the condition below the summation symbol. Using (1) and (2) we can transform this recurrence relation for $p_i(f)$ into a recurrence relation for $\Delta_i(w)$. We give only the resulting formula in Theorem 2.6. The complete derivation can be found in [7].

Definition 2.4. (a) For every $r \in \mathscr{B}_k$, let

$$S_\alpha(r) = (\tfrac{1}{2})^k \sum_{x \in \mathscr{B}_k} (-1)^{\alpha(x) \oplus \langle r, x \rangle}.$$

(b) For every w and $\tilde{v}$ let

$$Q_\alpha(\tilde{v}, w) = \begin{cases} \prod_{\substack{a \in M \\ w(a)=1}} S_\alpha(\tilde{v}(a)) & \text{if } v_j \leqslant w \text{ for } j = 1, \ldots, k, \\ 0 & \text{otherwise.} \end{cases}$$

To get some necessary estimates of the numbers $S_\alpha(r)$ we shall use the following facts. For a detailed proof, see [7, Lemma 4.7]. Let r denote an arbitrary element of $\mathscr{B}_k$.

Lemma 2.5. (a) $\sum_r S_\alpha(r)^2 = 1$.

(b) *α is balanced if and only if* $S_\alpha(0_k) = 0$.

(c) *α is nonlinear if and only if for all r we have* $|S_\alpha(r)| < 1$.

(d) *If α is nonlinear then* $\sum_{|r|=1} S_\alpha(r)^2 < 1$.

Theorem 2.6. *For all w and $i \geqslant 0$ we have*

$$\Delta_{i+1}(w) = \sum_{v_1, \ldots, v_k \leqslant w} Q_\alpha(\tilde{v}, w) \Delta_i(v_1) \ldots \Delta_i(v_k). \tag{4}$$

Since $\Delta_{i+1}(w)$ only depends on $\Delta_i(v)$ for $v \leqslant w$ in the formula (4), we can study the asymptotic behaviour of the sequence $\Delta_i(w)$ for $i \to \infty$ by induction on $|w|$. For this purpose we shall distinguish in formula (4) the terms containing at least one factor $\Delta_i(v)$, where v satisfies $0_M < v < w$ and the other terms containing only several occurrences of $\Delta_i(w)$ and $\Delta_i(0_M)$. The former terms will be estimated using the induction hypothesis. The latter terms, since $\Delta_i(0_M) = 1$, form a polynomial in $\Delta_i(w)$ with constant coefficients. For the separation of these two types of terms in (4), we shall use the following notation.

Definition 2.7. (a) For every $m \geqslant 1$ and $j \geqslant 0$, let $T_j(m) = \sum_{|r|=j} S_\alpha(r)^m$. For an arbitrary nonzero function w we shall abbreviate $T_j(|w|)$ as $T_j(w)$.

(b) $H_w = \{\tilde{v} \in \mathscr{B}_M^k \mid v_j \leqslant w$ for all $j = 1, \ldots, k$ and there exists j such that $v_j \notin \{0_M, w\}\}$.

Theorem 2.8. *If α is balanced then for every nonzero w and every $i \geqslant 0$ we have*

$$\Delta_{i+1}(w) = \sum_{j=1}^{k} T_j(w)\Delta_i(w)^j + \sum_{H_w} Q_\alpha(\tilde{v}, w)\Delta_i(v_1)\dots\Delta_i(v_k). \tag{5}$$

Proof. We shall prove (5) on the base of (4). By the definition of H_w, it is sufficient to prove that the first sum in (5) is equal to the sum of those terms of (4) satisfying $v_j \in \{0_M, w\}$ for all $j = 1, \dots, k$. Since $w \neq 0_M$, these terms are in one to one correspondence with $\mathscr{B}_k$, the term corresponding to a particular $r \in \mathscr{B}_k$ being determined as follows: $v_j = w \Leftrightarrow r_j = 1$. For the term corresponding to some $r \in \mathscr{B}_k$ and for $a \in M$ satisfying $w(a) = 1$, we obtain $\tilde{v}(a) = r$. Hence,

$$Q_\alpha(\tilde{v}, w)\Delta_i(v_1)\dots\Delta_i(v_k) = S_\alpha(r)^{|w|}\Delta_i(w)^{|r|}.$$

Since α is balanced and $|w| > 0$, the term corresponding to $r = 0_k$ is zero. The other terms form the sum

$$\sum_{j=1}^{k}\sum_{|r|=j} S_\alpha(r)^{|w|}\Delta_i(w)^j = \sum_{j=1}^{k} T_j(w)\Delta_i(w)^j. \qquad \square$$

If $\Delta_i(w)$ is small, then the greatest term in the first sum in (5) is its least degree term, i.e., the term with the minimal j such that $T_j(w) \neq 0$. If $w \neq 0$ is such that $\Delta_i(v) = 0$ for all $0_M < v < w$ and $\Delta_i(w) \neq 0$ for all $i \geqslant 0$, then this least degree term determines the rate of convergence of $\Delta_i(w)$ to zero. To get estimate of the rate of convergence of $\Delta_i(w)$ for an arbitrary w, we shall introduce the following notion.

Definition 2.9. Let $\deg(\alpha)$, the degree of a connective α, be the minimal weight of $r \in \mathscr{B}_k$ satisfying $S_\alpha(r) \neq 0$.

Notice that for every nonzero w the following holds. If $j < \deg(\alpha)$ then $T_j(w) = 0$. On the other hand, if j is equal to the degrees and $|w|$ is even then $T_j(w) > 0$.

3. A combinatorial characterization of the degree

We shall characterize the degree of α in a more transparent way. Let r, s, t, x denote elements of $\mathscr{B}_k$ throughout this section.

Theorem 3.1. *The degree of α is equal to the minimal number of the arguments of α which must be set to 0 or 1 in order to obtain a nonbalanced connective (as a subfunction) on the remaining arguments.*

Proof. Let $s, r \in \mathscr{B}_k$ be such that $s \leqslant r$. Let α' be the $(k - |r|)$-ary connective arisen from $\alpha(x)$ by setting x_j to s_j if $r_j = 1$. We shall call α' the s, r-restriction of α. Since the

domain of α' consists of $x \in \mathscr{B}_k$ satisfying $x \wedge r = s$, the s,r-restriction of α is balanced if and only if

$$\sum_{x \wedge r = s} (-1)^{\alpha(x)} = 0,$$

where $\wedge$ is applied to all coordinates of its arguments. For a fixed $r \in \mathscr{B}_k$, we shall rearrange the formula defining $S_\alpha(t)$ for an arbitrary $t \leqslant r$ using the following two facts. First, the sum over all $x \in \mathscr{B}_k$ can be divided into the sum over all $s \leqslant r$ of the sums over all x satisfying $x \wedge r = s$. Second, if $x \wedge r = s$ and $t \leqslant r$ then $\langle t, x\rangle = \langle t, s\rangle$. We obtain

$$S_\alpha(t) = (\tfrac{1}{2})^k \sum_{s \leqslant r} (-1)^{\langle t,s\rangle} \sum_{x \wedge r = s} (-1)^{\alpha(x)}.$$

If $r \in \mathscr{B}_k$ is fixed, the numbers $(-1)^{\langle t,s\rangle}$ for all $t \leqslant r$ and $s \leqslant r$ form a $2^{|r|}$ by $2^{|r|}$ matrix, which is regular, since it is a Hadamard matrix (Sylvester matrix in particular, see [5]). Consequently, the following two statements are equivalent:

(a) $S_\alpha(t) = 0$ for all $t \leqslant r$;

(b) the s,r-restriction of α is balanced for all $s \leqslant r$.

Hence, the minimal weight of r satisfying $S_\alpha(r) \neq 0$ is equal to the minimal weight of r, for which such a vector s exists that the s, r-restriction of α is not balanced. □

We shall present two examples of connectives of an arbitrary given degree. For the first one we need the following property of the balanced connectives.

Lemma 3.2. *Let $\alpha_1 = \alpha_2 \oplus \alpha_3$ and let α_2 and α_3 depend on disjoint subsets of variables of α_1. Then α_1 is balanced if and only if at least one of the two connectives α_2 and α_3 is balanced.*

Proof. Let $\tilde{t}$ be a random assignment of variables of α_1. Clearly, α_j is balanced if and only if

$$E[(-1)^{\alpha_j(\tilde{t})}] = 0.$$

Since the expected value of the product of two independent random variables is the product of the expected values of the two random variables, we have

$$E[(-1)^{\alpha_j(\tilde{t})}] = E[(-1)^{\alpha_2(\tilde{t}) \oplus \alpha_3(\tilde{t})}] = E(-1)^{\alpha_2(\tilde{t})}] \cdot E[(-1)^{\alpha_3(\tilde{t})}].$$

The desired result is a simple consequence. □

Theorem 3.3. *Let α_j for $j = 1, \ldots, s$ be nonconstant connectives depending on pairwise disjoint sets of variables. If $1 \leqslant d \leqslant s$ and α_j is balanced for $j = 1, \ldots, d$, then the degree of $\alpha_1 \oplus \cdots \oplus \alpha_s$ is at least d.*

Proof. We shall use the characterization of the degree through the setting variables to constants from Theorem 3.1. Since α_j for $j = 1, \ldots, d$ are nonconstant connectives, any of them depends at least on one variable. These variables are pairwise distinct. If we set any $d - 1$ variables to constants, there is some $1 \leqslant j \leqslant d$ such that α_j is left unchanged by this setting. Hence, $\alpha_1 \oplus \cdots \oplus \alpha_s$ is balanced by Lemma 3.2. It follows that $\deg(\alpha)$ is greater than $d - 1$. □

Example 3.4. *The degree of* $x_1x_2 \oplus x_3 \oplus \cdots \oplus x_d$ *is* $d - 2$.

Proof. Since the variables are balanced connectives, the degree is at least $d - 2$ by the last theorem. On the other hand, any setting of variables $x_3, \ldots, x_d$ to any constants leads to either x_1x_2 or $x_1x_2 \oplus 1$. Both of these connectives are not balanced. Hence, the degree is at most $d - 2$. □

The following examples shows that the combining functions by addition modulo 2 is not the only possible way to construct a connective of an arbitrarily high degree. Recall that sel, the selection function, satisfies: $\mathrm{sel}(0, y, z) = y$ and $\mathrm{sel}(1, y, z) = z$.

Example 3.5. *Let* $d \geqslant 1$ *and* $\alpha = \mathrm{sel}(x, y_1 \oplus \cdots \oplus y_d, z_1 \oplus \cdots \oplus z_d)$. *The degree of* α *is* d. *Moreover, if* $\alpha = \alpha_1 \oplus \alpha_2$ *and both* α_1 *and* α_2 *depend on disjoint sets of variables, then one of these two connectives is identically equal to a constant.*

Proof. If we set $y_1 = \cdots = y_d = 0$, the connective α is changed to $\mathrm{sel}(x, 0, z_1 \oplus \cdots \oplus z_d)$. It easy to see that this connective takes the value 1 with the probability 1/4 on a random assignment of its variables. Hence, it is not a balanced connective and the degree of α is at most d by Theorem 3.1.

On the other hand, let $\mathscr{P} \subseteq \{x, y_1, \ldots, y_d, z_1, \ldots, z_d\}$ and $|\mathscr{P}| < d$. Let $\tilde{t}$ be a random assignment of variables of α such that all variables in $\mathscr{P}$ are set to some constants and all the remaining variables are set to independent random values. Let $\alpha_y = y_1 \oplus \cdots \oplus y_d$ and $\alpha_z = z_1 \oplus \cdots \oplus z_d$. Since $|\mathscr{P}| < d$, both $\alpha_y(\tilde{t})$ and $\alpha_z(\tilde{t})$ take the value 1 with the probability $\frac{1}{2}$. Let $x(\tilde{t})$ be the value of the variable x in the assignment $\tilde{t}$. Since $x(\tilde{t}), \alpha_y(\tilde{t})$ and $\alpha_z(\tilde{t})$ are mutually independent, we obtain

$$P(\alpha(\tilde{t}) = 1) = P(\alpha_y(\tilde{t}) = 1)P(x(\tilde{t}) = 0) + P(\alpha_z(\tilde{t}) = 1)P(x(\tilde{t}) = 1)$$

$$= \tfrac{1}{2}(P(x(\tilde{t}) = 0) + P(x(\tilde{t}) = 1)) = \tfrac{1}{2}.$$

It follows that any setting of variables in $\mathscr{P}$ to constants transforms α into a balanced connective. Hence, the degree of α is at least d.

Let $\alpha = \alpha_1 \oplus \alpha_2$, where α_1 and α_2 depend on disjoint sets of variables of α. Let $\mathscr{P}_i$ be the set of variables α_i depends on. Without loss of generality, we may assume that $x \in \mathscr{P}_1$.

Let us assume for a moment that $z_j \in \mathscr{P}_2$ for some $j = 1, \ldots, d$. In this case, α_1 does not depend on z_j. Let us set all the variables y_l for all $l \leqslant d$ and z_l for all $l \neq j$ to zero. By this setting, α is converted to xz_j, while α_1 and α_2 are converted to connectives depending only on one variable, x and z_j, respectively. Hence, $\alpha_1 \oplus \alpha_2$ is converted to $c_1 x \oplus c_2 z_j \oplus c_3$ for appropriate constants $c_1, c_2, c_3 \in \{0, 1\}$. Since xz_j is not a linear connective, it is a contradiction.

By a similar way, we can disprove any of the hypothesis $y_j \in \mathscr{P}_2$ for any $j = 1, \ldots, d$. In this case, we set all the variables with the exception of x and y_j to zero. Consequently, α is converted to $(\neg x)y_j$ and $\alpha_1 \oplus \alpha_2$ to a linear function as in the previous case. Again, it is a contradiction. Hence, $\mathscr{P}_2$ is empty and α_2 is a constant. □

4. Asymptotic estimate for an arbitrary degree connective

Let M, G and α be the same as in Section 2. All of the computations of the asymptotic properties of p_i are done using Δ_i. On the other hand, the results and the assumptions of the theorems are more understandable in terms of the probabilities of some events depending on $\tilde{g}_i$. Hence, some of the properties of the distribution p_i, which are relevant to our considerations, are formulated in two equivalent ways and we establish the translation between the language of probabilities and the language of Δ_i.

We shall often use the restriction of $\tilde{g}_i$ on subsets of M. If $A \subseteq M$, then we denote the restriction of $\tilde{g}_i$ on A as $\tilde{g}_i|_A$.

The distribution of $\tilde{g}_i|_A$ is uniform if and only if the events $\tilde{g}_i(a) = 1$ are independent for all $a \in A$ and have the probability $\frac{1}{2}$. Let us assume that $\tilde{g}_i$ satisfies this property. Recall the fact that $\tilde{g}_{i+1}$ arise from k independent copies of $\tilde{g}_i$ by combining via α. Since α is a balanced connective, the probability of $\tilde{g}_{i+1}(a) = 1$ is again equal to $\frac{1}{2}$ for every $a \in A$. Moreover, since the value of $\tilde{g}_{i+1}(a)$ for a given $a \in A$ only depends on the values of $\tilde{g}_{i,j}(a)$ for $j = 1, \ldots, k$ and these k-tuples of values are independent for all $a \in A$, the values $\tilde{g}_{i+1}(a)$ are again independent for all $a \in A$.

It follows that if the distribution of $\tilde{g}_0|_A$ is uniform on $\mathscr{B}_A$, so is the distribution of $\tilde{g}_i|_A$ for all $i \geqslant 0$.

Definition 4.1. Let $\mathrm{rank}(G)$ be the size of the smallest nonempty $A \subseteq M$ such that the distribution of $\tilde{g}_0|_A$ is not uniform. If $\tilde{g}_0|_A$ has the uniform distribution on $\mathscr{B}_A$ for all nonempty subsets A of M, $\mathrm{rank}(G)$ is undefined.

It can easily be seen from the relations defining the Fourier transform that $\tilde{g}_i$ has the uniform distribution if and only if for all nonzero functions w, $\Delta_i(w) = 0$. Analogously, the restriction $\tilde{g}_i|_A$ has the uniform distribution on $\mathscr{B}_A$ if and only if $\Delta_i(w) = 0$ for all nonzero functions w which are zero everywhere outside A. The following lemma follows from this and from the fact that the uniform distribution is preserved by combining via α.

Lemma 4.2. (i) $\mathrm{rank}(G)$ *is equal to the minimal* $|w|$, *where* $w \neq 0_M$ *and* $\Delta_0(w) \neq 0$.
(ii) *For every* w *such that* $0 < |w| < \mathrm{rank}(G)$, *and for every* $i \geqslant 0$ $\Delta_i(w) = 0$.

We get a consequence that if $|w| = \mathrm{rank}(G)$ then the second sum in the recurrence formula (5) can be dropped and the following formula takes place in this situation:

$$\Delta_{i+1}(w) = \sum_{j=1}^{k} T_j(w)\Delta_i(w)^j. \tag{6}$$

Any sufficient condition for the convergence of p_i to the uniform distribution should imply $\Delta_i(w) \to 0$ for all nonzero w and hence, in particular, for $|w| = \mathrm{rank}(G)$. If $|w| = 1$ then the recurrence relation (6) may have other fixed points besides $\Delta_i(w) = 0$, and the fixed point $\Delta_i(w) = 0$ may not be stable for some α. For these reasons, we have no simple sufficient condition for the convergence of p_i to the uniform distribution if $\mathrm{rank}(G) = 1$.

Hence, we always assume $\mathrm{rank}(G) \geqslant 2$. Notice that this condition is equivalent to the following one: for every $a \in M$ $P(\tilde{g}_0(a) = 1) = \frac{1}{2}$. The starting set G satisfying this property will be called unbiased.

Definition 4.3. The starting set G will be called degenerate, if there are $a, b \in M$, $a \neq b$ such that either $f(a) = f(b)$ for all $f \in G$ or $f(a) \neq f(b)$ for all $f \in G$.

If $f(a) = f(b)$ for all $f \in G$ then, of course, combining functions in G via any connective cannot lead to functions satisfying $f(a) \neq f(b)$. On the other hand, if $f(a) \neq f(b)$ for all $f \in G$ then, for many connectives α, the set of all combinations of functions in G via α contains both the functions satisfying $f(a) = f(b)$ and the functions satisfying $f(a) \neq f(b)$. However, if α is self-dual (i.e α satisfies $\alpha(\neg t) = \neg\alpha(t)$ for all $t \in \mathscr{B}_k$) or α satisfies $\alpha(\neg t) = \alpha(t)$ for all $t \in \mathscr{B}_k$ then the set of all combinations of functions from G via α is again degenerate. In order to obtain results for connectives including self-dual connectives and the connectives satisfying $\alpha(\neg t) = \alpha(t)$, we restrict ourselves to nondegenerate G.

The condition that G is unbiased and nondegenerate, is sufficient condition for the convergence of p_i to the uniform distribution, which applies to all balanced nonlinear connectives. This follows from Lemma 4.6 below.

The properties of being unbiased and nondegenerate can simply be reformulated in terms of $\Delta_0(w)$.

Lemma 4.4. *The starting set* G *is*
(i) *unbiased if and only if for every function* w *of weight* $|w| = 1$, $\Delta_0(w) = 0$,
(ii) *nondegenerate if and only if for every function* w *of weight* $|w| = 2$, $\Delta_0(w) < 1$.

Proof. Statement (i) follows from Lemma 4.2. If $|w| = 2$, then there exist elements $a, b \in M$ such that $w(a) = w(b) = 1$ and for any other element $c \in M$, $w(c) = 0$. Since, $\Delta_0(w) = E[(-1)^{\tilde{g}_0(a) \oplus \tilde{g}_0(b)}]$, the statement (ii) follows. □

Notice that Lemma 4.4 allows us to generalize the properties of being unbiased and nondegenerate from starting sets to an arbitrary distribution p_0 and $\mathscr{B}_M$. In fact, all the proofs of the present section apply to this more general case as well.

Now, we shall continue transforming the recurrence formula for $\Delta_i(w)$ into the form suitable for deriving asymptotic estimates. We shall use the formula (5) and all the notation and assumptions applying to it.

Lemma 4.5. *Let $w \neq 0$ and let for all $i \geqslant 0$,*

$$b_i = \sum_{j=1}^{k} T_j(w)\Delta_i(w)^{j-1},$$

$$h_i = \sum_{H_w} Q_\alpha(\tilde{v}, w)\Delta_i(v_1)\cdots\Delta_i(v_k).$$

Then for all $i \geqslant 0$,

$$\Delta_i(w) = \sum_{j=0}^{i-1} h_j \cdot \prod_{m=j+1}^{i-1} b_m + \Delta_0(w)\cdot \prod_{m=0}^{i-1} b_m$$

is satisfied.

Proof. By comparing the formulas defining b_i and h_i with (5), we can see that the following holds for all $i \geqslant 0$: $\Delta_{i+1}(w) = b_i\Delta_i(w) + h_i$. The statement of the lemma follows from this by a simple induction argument. □

Lemma 4.6. *Let α be balanced and nonlinear. Let G be unbiased and nondegenerate. Then there exists positive $c < 1$ such that $\Delta_i(w) = O(c^i)$ for all nonzero $w \in \mathscr{B}_M$ and $i \to \infty$.*

Proof. Since G is unbiased, $\operatorname{rank}(G)$ is at least 2. Hence, if $|w| = 1$, then $\Delta_i(w) = 0$ for all $i \geqslant 0$ by Lemma 4.2 and there is nothing to prove.

Further, we shall proceed by induction on $|w|$. Let $|w| \geqslant 2$ and let the statement be true for all v satisfying $0_M < v < w$. Then there exists $a < 1$ such that $\Delta_i(v) = O(a^i)$ simultaneously for all v in consideration.

We shall derive an upper bound on the absolute value of the quantities h_i and b_i appearing in the previous lemma. Every term in the sum defining h_i contains at least one term which is $O(a^i)$. Since $|\Delta_i(v)| \leqslant 1$ for all v and there is only a finite number of terms in the sum, $h_i = O(a^i)$.

In the next step, we show that there exists $b < 1$ such that $|b_i| < b$ for all $i \geqslant 0$. There are two cases to be considered separately.

Case 1: $|w| = 2$. If $\operatorname{rank}(G) > 2$ then $\Delta_i(w) = 0$ for all $i \geqslant 0$ by Lemma 4.2. Let us assume $\operatorname{rank}(G) = 2$. Since $|w|$ is even, $T_j(w) \geqslant 0$ for all $j = 1, \ldots, k$ and by

Lemma 2.5(a), $\sum_{j=1}^{k} T_j(w) = 1$. By the properties of the Fourier transform, $|\Delta_i(w)| \leqslant 1$ for all $i \geqslant 0$. It follows that $|b_i| \leqslant 1$ for all $i \geqslant 0$. In fact, $h_i = 0$ holds for all $i \geqslant 0$ in this case and then $|\Delta_i(w)| \leqslant |\Delta_0(w)|$ holds for all $i \geqslant 0$ by Lemma 4.5.

Let $b = \sum_{j=1}^{k} T_j(w)|\Delta_0(w)|^{j-1}$. Clearly, $|b_i| \leqslant b$ for all $i \geqslant 0$. It remains to show that $b < 1$. Since G is nondegenerate, we have $|\Delta_0(w)| < 1$, and by Lemma 2.5(d) we get $T_1(w) < 1$. The computation can be finished as follows: $b \leqslant T_1(w) + |\Delta_0(w)| \sum_{j=2}^{k} T_j(w) \leqslant T_1(w) + |\Delta_0(w)|(1 - T_1(w)) < T_1(w) + (1 - T_1(w)) = 1$.

Case 2: $|w| \geqslant 3$. Let

$$b = \sum_{j=1}^{k} |T_j(w)|.$$

Clearly, $|b_i| \leqslant b$. Since α is nonlinear, there is at least one $r \in \mathscr{B}_k$ such that $|S_\alpha(r)| \in (0, 1)$. Hence, by definition of $T_j(w)$,

$$b = \sum_{j=1}^{k} |T_j(w)| < \sum_{j=1}^{k} T_j(2) = 1.$$

Now, we apply the bounds $h_i = O(a^i)$ and $|b_i| \leqslant b < 1$ to the bound proved in Lemma 4.5 and we obtain the following:

$$\Delta_i(w) = O\left(\sum_{j=0}^{i-1} b^{i-j} a^j\right) + O(b^i).$$

Let $c < 1$ be such that $b < c$ and $a \leqslant c$. Then

$$\Delta_i(w) = O\left(c^i \sum_{j=0}^{i-1} (b/c)^{i-j}\right) + O(c^i) = O(c^i).$$

The induction step is complete. □

Theorem 4.7. *Let α be balanced and nonlinear and let $d = \deg(\alpha)$. Let G be unbiased and nondegenerate. Then there exists a positive constant $c < 1$ such that for every function $f \in \mathscr{B}_M$ the following holds:*

$$p_i(f) - (\tfrac{1}{2})^{|M|} = \begin{cases} O(c^i) & \text{if } d = 1, \\ O(c^{d^i}) & \text{if } d \geqslant 2. \end{cases}$$

Proof. By Lemma 4.6, for every nonzero $w \in \mathscr{B}_M$, there is a positive constant $a < 1$ such that $\Delta_i(w) = O(a^i)$. There is only a finite number of the functions w. Hence, if c is the maximum of these numbers a, the corresponding estimate applies simultaneously to all nonzero w. If $d = 1$, the proof can be finished by applying the upper bound from (3).

If $d \geqslant 2$, we shall use the formula (4). Let q_i be the maximum of $|\Delta_i(w)|$ for all nonzero w. Clearly, $q_i \leqslant 1$ for all $i \geqslant 0$. By the first part of the proof, q_i tends to 0 if i tends to infinity. Since $|Q_\alpha(\tilde{v}, w)| \leqslant 1$ for all $\tilde{v}$ and w, the absolute value of an arbitrary

term in the formula (4) corresponding to the k-tuple $[v_1, \ldots, v_k]$. Notice that if $l < d$ then $S_\alpha(\tilde{v}(x)) = 0$ for all $x \in M$. Hence, if $l < d$ and $w \neq 0_M$ then $Q_\alpha(\tilde{v}, w) = 0$. It follows that for every nonzero w and every $i \geqslant 0$ we have $|\Delta_{i+1}(w)| \leqslant mq_i^d$, where m is the number of the terms in (4). Hence, $q_{i+1} \leqslant mq_i^d$ for every $i \geqslant 0$. The last formula can be transformed into

$$m^{1/(d-1)}q_{i+1} \leqslant (m^{1/(d-1)}q_i)^d.$$

Since $q_i \to 0$, there exists i_0 such that for

$$b = m^{1/(d-1)}q_{i_0},$$

we have $b < 1$. It follows by induction on i that if $i \geqslant i_0$, then $m^{1/(d-1)}q_i \leqslant b^{d^{i-i_0}}$. Hence, if $c = b^{d^{-i_0}}$, we obtain $q_i = O(c^{d^i})$. This estimate holds for $\Delta_i(w)$ for all nonzero w. Using (3) again, we obtain the desired result. □

Let us denote

$$\phi_d(i) = \begin{cases} i, & d = 1, \\ d^i, & d \geqslant 2. \end{cases}$$

The logarithm of the upper bound in Theorem 4.7 is a function of the type $-\Theta(\phi_d(i))$, where $d = \deg(\alpha)$. Moreover, the classes of functions $e^{-\Theta(\phi_d(i))}$ form a strictly decreasing hierarchy of magnitude for $d = 1, 2, 3, \ldots$ etc. In the following theorem we show that under a little stronger assumptions on G than before, we may obtain a lower bound on $\max_{f \in \mathscr{B}_M} |p_i(f) - (\frac{1}{2})^{|M|}|$, which belongs to the same level of this hierarchy as the corresponding upper bound. Hence, the classification of the connectives according to the degree yields also a classification according to the asymptotic rate of convergence of p_i to the uniform distribution.

Theorem 4.8. *Let α be balanced and nonlinear connective and let $d = \deg(\alpha)$. Let the starting set G be unbiased and nondegenerate. Let* $\mathrm{rank}(G)$ *be even and there exists $w \in \mathscr{B}_M$ such that* $|w| = \mathrm{rank}(G)$ *and $\Delta_0(w) > 0$. Then*

$$\max_{f \in \mathscr{B}_M} |p_i(f) - (\tfrac{1}{2})^{|M|}| = \begin{cases} e^{-\Theta(i)} & \text{if } d = 1, \\ e^{-\Theta(d^i)} & \text{if } d \geqslant 2. \end{cases}$$

Proof. Let $\phi_d(i)$ be as above. By the definition of Θ, the statement of the theorem is equivalent to

$$e^{-c_2\phi_d(i)} \leqslant \max_{f \in \mathscr{B}_M} |p_i(f) - (\tfrac{1}{2})^{|M|}| \leqslant e^{-c_1\phi_d(i)}$$

for every sufficiently large i, where c_1, c_2 are some positive constants. The upper bound has been proved already in Theorem 4.7.

Let us turn to the lower bound. Let $|w| = \mathrm{rank}(G)$ and $\Delta_0(w) > 0$. The sequence $\Delta_i(w)$ satisfies the relation (6), where $T_j(w) \geqslant 0$ for all $j = 1, \ldots, k$ and $T_d(w)$ is positive. It follows that if $d = 1$ then

$$\Delta_{i+1}(w) \geqslant T_1(w)\Delta_i(w)$$

and if $d \geqslant 2$ then

$$T_d(w)^{1/(d-1)}\Delta_{i+1}(w) \geqslant (T_d(w)^{1/(d-1)}\Delta_i(w))^d.$$

In the former case we obtain

$$\Delta_i(w) \geqslant \mathrm{e}^{\ln T_1(w) i + \ln \Delta_0(w)}$$

and in the latter one we obtain

$$T_d(w)^{1/(d-1)}\Delta_i(w) \geqslant \mathrm{e}^{(1/(d-1))\ln T_d(w) + \ln \Delta_0(w))d^i}.$$

In either case we obtain

$$\Delta_i(w) \geqslant \mathrm{e}^{-c_2\phi_d(i)}$$

for some real number $c_2 > 0$. Hence, by inequality (3),

$$\max_{f \in \mathscr{B}_M} |p_i(f) - (\tfrac{1}{2})^{|M|}| \geqslant \mathrm{e}^{-c_2\phi_d(i) - |M|\ln 2} \geqslant \mathrm{e}^{-c_3\phi_d(i)}$$

for some $c_3 > c_2$ and sufficiently large i. □

5. The connectives of degree one

For the connectives of degree one we shall prove a stronger result than that contained in Theorem 4.8. Namely, we shall compute the leading term of $p_i(f) - (\frac{1}{2})^{|M|}$ (Theorem 5.5 below). Let α be a nonlinear k-ary connective of degree 1.

Let rank be a shorthand of rank(G). We assume throughout this section that rank is even. This assumption guarantees that $T_1(\mathrm{rank}) > 0$. Then, we shall prove that for $|w| = \mathrm{rank}$ we have $\Delta_i(w) = \sigma(\Delta_0(w))T_1(\mathrm{rank})^i + O(T_1(\mathrm{rank})^{2i})$, where σ is a real function, depending on rank and α. For $|w| > \mathrm{rank}$ we shall prove that $\Delta_i(w) = O(U^i)$, where $0 < U < T_1(\mathrm{rank})$. Notice that by Lemma 2.5(d) $T_1(\mathrm{rank}) < 1$. Let us consider the expansion (2) of $p_i(f)$. It follows that for large i, the terms satisfying $|w| = \mathrm{rank}$ are the leading terms of $p_i(f) - (\frac{1}{2})^{|M|}$, provided there is at least one w with $|w| = \mathrm{rank}$ and $\sigma(\Delta_0(w)) \neq 0$ and the terms corresponding to these w do not cancel out.

The constant $\sigma(\Delta_0(w))$ is the limit of an infinite iterative process determined by $\Delta_0(w)$ and relation (6). By simple reasons, $\sigma(0) = 0$, but for a nonzero Δ we are not able to express $\sigma(\Delta)$ explicitly. If $\Delta < 0$, it makes troubles even to establish whether $\sigma(\Delta) \neq 0$ or not. Hence, it may be hard to use the asymptotic formula derived here in general case. We shall use it in the end of this section in the case that G is a linear code.

If $0 < |w| <$ rank, then Lemma 4.2 implies $\Delta_i(w) = 0$ for all $i \geqslant 0$. If $|w| =$ rank and $\Delta_0(w) = 0$ then the same follows by induction from (6). The asymptotic estimate of $\Delta_i(w)$ in the remaining cases will be done using the following technical lemma.

Lemma 5.1. *Let x_i, h_i, b_i be arbitrary sequences of real numbers satisfying $x_{i+1} = b_i x_i + h_i$ for all $i \geqslant 0$. Let $h_i = O(a^i)$ for some real number a and let the sum of $|b_i - b|$ over all $i \geqslant 0$ be convergent for some real number b. Let $|b| \leqslant c$ and $a < c$. Then $x_i = O(c^i)$.*

Proof. By a simple induction argument the following holds for all $i \geqslant 0$:

$$x_i = \sum_{j=0}^{i-1} h_j \cdot \prod_{m=j+1}^{i-1} b_m + x_0 \cdot \prod_{m=0}^{i-1} b_m.$$

Since $|b_m| \leqslant c + |b_m - b|$, the products of b_m may be bounded for every $i \geqslant j \geqslant 0$ as follows:

$$\left|\prod_{m=j}^{i-1} b_m\right| \leqslant c^{i-j} \prod_{m=j}^{i-1} \left(1 + \frac{|b_m - b|}{c}\right) \leqslant c^{i-j} \exp\left(\sum_{m=0}^{\infty} \frac{|b_m - b|}{c}\right).$$

By combining these two facts, we obtain

$$x_i = O\left(\sum_{j=0}^{i-1} c^{i-j} a^j\right) + O(c^i).$$

It means that

$$x_i = O\left(c^i \sum_{j=0}^{\infty} (a/c)^j\right) + O(c^i) = O(c^i). \qquad \square$$

Lemma 5.2. *Let α be balanced and nonlinear. Let G be unbiased, nondegenerate and let* rank(G) *be even. Then there exists a real function σ such that for every w, $|w| =$* rank *we have*

(i) $\Delta_i(w) = \sigma(\Delta_0(w)) T_1(\text{rank})^i + O(T_1(\text{rank})^{2i})$

(ii) *if, in addition, $\Delta_0(w) \geqslant 0$ then $\sigma(\Delta_0(w)) \geqslant \Delta_0(w)$.*

Proof. Since rank is even, $1 > T_1(\text{rank}) > 0$ and $T_j(\text{rank}) \geqslant 0$ for all $j = 2, \ldots, k$, by definition of $T_j(m)$ and by Lemma 2.5. Moreover, let $|w| =$ rank. Then (6) takes place.

By Theorem 4.7, $\Delta_i(w) = O(a^i)$ for some $0 < a < 1$. Hence, the sum of $|\sum_{j=1}^{k} T_j(\text{rank}) \Delta_i(w)^{j-1} - T_1(w)|$ over all $i \geqslant 0$ is convergent. It follows that the recurrence relation (6) satisfies the assumptions of Lemma 5.1 with $h_i = 0$, $x_i = \Delta_i(w)$ and $b = c = T_1(\text{rank})$. We obtain $\Delta_i(w) = O(T_1(\text{rank})^i)$.

Let $z_i = \Delta_i(w) T_1(\text{rank})^{-i}$. Using relation (6) we obtain

$$z_{i+1} - z_i = T_1(\text{rank})^{-i-1} \sum_{j=2}^{k} T_j(\text{rank}) \Delta_i(w)^j = O(T_1(\text{rank})^i).$$

Hence, the sequence z_i is convergent. Let us denote its limit as $\sigma(\Delta_0(w))$. Clearly,

$$z_i = \sigma(\Delta_0(w)) - \sum_{m=i}^{\infty} (z_{m+1} - z_m) = \sigma(\Delta_0(w)) + O(T_1(\text{rank})^i).$$

The proof of (i) may be finished using the definition of z_i.

If $\Delta_0(w) \geqslant 0$, relation (6) implies $\Delta_{i+1}(w) \geqslant \Delta_i(w) T_1(\text{rank})$. Hence, $\Delta_i(w) \geqslant \Delta_0(w) T_1(\text{rank})^i$. This and the part (i) of this lemma imply (ii). □

In order to show that the terms $\Delta_i(w)(-1)^{\langle w, f\rangle}$ for $|w| = \text{rank}$ are the leading terms of the expansion of $p_i(f)$, we need the following lemma.

Lemma 5.3. *If* $U \geqslant \max(|T_1(\text{rank}+1)|, T_1(\text{rank}+2))$ *and* $T_1(\text{rank}) > U > T_1(\text{rank})^2$ *then for all nonzero functions* w *with the weight at least* rank + 1 *we have* $\Delta_i(w) = O(U^i)$. *Moreover, the number* U *with these properties exists.*

Proof. The number $T_1(m)$ is a sum of $S_\alpha(r)^m$ for $r \in \mathscr{B}_k$ and $|r| = 1$. Since $\deg(\alpha) = 1$ and α is nonlinear, $0 < \max_{|r|=1} |S_\alpha(r)| < 1$. Hence, since rank is even, if rank $< m$ then $T_1(\text{rank}) > |T_1(m)|$. Analogously, if rank + 2 $\leqslant m$ then $T_1(\text{rank}+2) \geqslant |T_1(m)|$. It follows that for all $|w| \geqslant \text{rank} + 1$:

$$T_1(\text{rank}) > \max(|T_1(\text{rank}+1)|, T_1(\text{rank}+2)) \geqslant T_1(w)|. \tag{7}$$

Since $1 > T_1(\text{rank}) > T_1(\text{rank})^2$, it implies, in particular, that a number U satisfying the assumptions of the lemma exists.

We shall prove the desired estimate of $\Delta_i(w)$ by induction on $|w|$. Let $|w| \geqslant \text{rank} + 1$ and let the induction hypothesis $\Delta_i(v) = O(U^i)$ hold for all v satisfying $|w| > |v| \geqslant \text{rank} + 1$. Notice that the induction hypothesis is empty if $|w| = \text{rank} + 1$. Let

$$H'_w = \{\tilde{v} \in H_w \mid v_j \neq w \text{ for } j = 1, \ldots, k\}$$

and

$$h'_i = \sum_{H'_w} Q_\alpha(\tilde{v}, w) \Delta_i(v_1) \ldots \Delta_i(v_k).$$

We shall transform (5) into (8) in such a way that all terms containing at least one occurrence of $\Delta_i(w)$ are collected together. These terms are all the terms in the first sum in (5) and those terms from the second one, which are not included into H'_w. We shall take the term $T_1(w)\Delta_i(w)$ without changes. The terms $T_j(w)\Delta_i(w)^j$ where $j \geqslant 2$ may be estimated as $O(\varepsilon^i)\Delta_i(w)$ for an appropriate $\varepsilon < 1$ by Lemma 4.6. All terms in the sum over H_w in (5), which are not in h'_i, contain a factor $\Delta_i(w)$ and also $\Delta_i(v)$ for some nonzero v different from w. By applying Lemma 4.6 again, we obtain that these terms are also $O(\varepsilon^i)\Delta_i(w)$ for an appropriate $\varepsilon < 1$. Getting these facts together, we obtain

$$\Delta_{i+1}(w) = (T_1(w) + O(\varepsilon^i))\Delta_i(w) + h'_i, \tag{8}$$

where $\varepsilon < 1$ is properly chosen.

Now, we shall derive an estimate of h'_i by estimating all terms in the sum defining it. Let $\tilde{v} \in H'_w$ and $Q_\alpha(\tilde{v}, w) \neq 0$. Then for all $x \in M$ satisfying $w(x) = 1$ we have $S_\alpha(\tilde{v}, w)) \neq 0$ and so $\tilde{v}(x) \neq 0_k$ for all these x. Since $v_j < w$ for all $j = 1, \ldots, k$, there are at least two nonzero functions in $\tilde{v}$. If the weight of one of them is less than rank then the corresponding term is equal to zero. If the weight of both is equal to rank then the whole term is $O(T_1(\text{rank})^{2i})$ by Lemma 5.2. If one of the two functions is of weight rank and the other of weight at least rank + 1 then the whole term is $O(T_1(\text{rank})^i U^i)$, because of the induction hypothesis. Analogously, if both functions are of weight at least rank + 1 then the term is $O(U^{2i})$. Since $U < T_1(\text{rank})$, it follows that every term in the sum defining h'_i is $O(T_1(\text{rank})^{2i})$ and hence so is h'_i.

Using this estimate of h'_i, we can see that the recurrence relation (8) satisfies the assumptions of Lemma 5.1 with $x_i = \Delta_i(w)$, $a = T_1(\text{rank})^2$, $b = T_1(w)$ and $c = U$, provided $|T_1(w)| \leqslant U$. This is true due to (7) and the assumptions of the lemma. Hence, we obtain $\Delta_i(w) = O(U^i)$. □

Under the assumption that rank $\geqslant 4$, we can simplify the assumptions of Lemma 5.3. Namely, we can take $U = \max(|T_1(\text{rank} + 1)|, T_1(\text{rank} + 2))$. In order to verify this, we apply the Cauchy inequality to the numbers $S_\alpha(r)^{\text{rank}/2+1}$ and $S_\alpha(r)^{\text{rank}/2-1}$ for all r satisfying $|r| = 1$. We obtain

$$\left(\sum_{|r|=1} S_\alpha(r)^{\text{rank}}\right)^2 \leqslant \left(\sum_{|r|=1} S_\alpha(r)^{\text{rank}+2}\right) \cdot \left(\sum_{|r|=1} S_\alpha(r)^{\text{rank}-2}\right).$$

Since rank $\geqslant 4$ and α is not linear, we can use Lemma 2.5 to prove that the rightmost sum in this inequality is strictly less than 1. Hence, by reading the inequality in terms of $T_1(m)$, we obtain $T_1(\text{rank})^2 < T_1(\text{rank} + 2) \leqslant U$.

Now, we shall express the leading term of $p_i(f) - (\frac{1}{2})^{|M|}$. This term is a geometric sequence with the quotient $T_1(\text{rank})$ and a coefficient which depends on the function f.

Definition 5.4. For all boolean functions f, let

$$\mathscr{D}(f) = \sum_{|w|=\text{rank}} \sigma(\Delta_0(w))(-1)^{\langle w, f\rangle}.$$

Theorem 5.5. *Let G be unbiased and nondegenerate and let α be a nonlinear connective of degree* 1. *Let $U < T_1(\text{rank})$ be as in Lemma* 5.3. *Then for every boolean function f we have*

$$p_i(f) = (\tfrac{1}{2})^{|M|}(1 + \mathscr{D}(f) T_1(\text{rank})^i + O(U^i)).$$

Proof. By the relation (2) we have

$$p_i(f) = (\tfrac{1}{2})^{|M|}\left(1 + \sum_{w \neq 0_M} \Delta_i(w)(-1)^{\langle w, f\rangle}\right).$$

Since $\Delta_i(w) = 0$ if $1 \leqslant |w| < \mathrm{rank}(G)$, the desired result may be obtained using Lemmas 5.2 and 5.3. □

In order to simplify the expression of $\mathscr{D}(f)$, we introduce one more assumption: G is a linear code over the two element field $\{0,1\}$, i.e. it is closed under $\oplus$. We shall prove that this assumption guarantees that $\Delta_0(w) \in \{0,1\}$ for all w. Hence, there is only one possible nonzero value of $\sigma(\Delta_0(w))$ for $|w| = \mathrm{rank}$, namely $\sigma(1)$.

If G is a linear code, the structure of Δ_0 is completely described by the dual code $G^\perp = \{w \mid (\forall f \in G)\langle f, w\rangle = 0\}$. This result follows from the treatment of the Fourier transform of general codes in [5, Ch. 5, Section 5]. For convenience, we give a complete proof.

Theorem 5.6. *Let $G \subseteq \mathscr{B}_M$. Then the following two statements hold:*

(a) *The statement G is a linear code if and only if for all functions w we have $\Delta_0(w) \in \{0,1\}$.*

(b) *If G is a linear code then for all functions w we have $\Delta_0(w) = 1$ if and only if $w \in G^\perp$.*

Proof. Let G be a linear code. If w is fixed then $\langle w, f\rangle$ is a linear mapping of $f \in G$ into $\{0,1\}$. There are two possibilities: either $\langle w, f\rangle = 0$ for all $f \in G$ or this is true for exactly one half of the functions $f \in G$. The former case takes place if $w \in G^\perp$ and the latter one otherwise. Hence,

$$\Delta_0(w) = \frac{1}{|G|}\sum_{f\in G}(-1)^{\langle w,f\rangle} = \begin{cases} 1 & \text{if } w \in G^\perp, \\ 0 & \text{if } w \notin G^\perp. \end{cases}$$

It proves (b) and the left to right implication of (a).

We shall prove the remaining direction of (a). Let us assume for a moment that $k = 2$ and $\alpha = \oplus$. Then, by definition, $\tilde{g}_1 = \tilde{g}_{0,1} \oplus \tilde{g}_{0,2}$. We could use Theorem 2.6 to obtain an expression for $\Delta_1(w)$, but in this particular case it may be done in a much simpler way. By the theorem on the expected value of the product of two independent real random variables, we obtain

$$\Delta_1(w) = E[(-1)^{\langle w, \tilde{g}_{0,1}\oplus\tilde{g}_{0,2}\rangle}] = E[(-1)^{\langle w,\tilde{g}_{0,1}\rangle}]\cdot E[(-1)^{\langle w,\tilde{g}_{0,2}\rangle}] = \Delta_0(w)^2$$

for all w. Hence, $\Delta_0(w) \in \{0,1\}$ for all w, then $\tilde{g}_0$ and $\tilde{g}_1$ have the same distribution. Clearly, it implies that G is closed under $\oplus$ and hence it is a linear code. □

Definition 5.7. Let $\mathscr{A}$ be the set of the minimum weight functions in $G^\perp$.

Notice that Theorem 5.6 implies that the weight of the functions in $\mathscr{A}$ is exactly $\mathrm{rank}(G)$. Moreover, if G is nondegenerate then $\Delta_0(w) = 0$ for $|w| = 2$ and hence $\mathrm{rank}(G) \geqslant 3$. Since we assume that rank is even, it is at least 4.

Hence, for linear codes we have

$$\mathscr{D}(f) = \sigma(1) \sum_{w \in \mathscr{A}} (-1)^{\langle w, f \rangle}.$$

It means that we can compute $\mathscr{D}(f)$ up to a positive multiplicative constant without computing the exact value of $\sigma(1)$. This is sufficient, for example, to study the functions f the probability $p_i(f)$ of which converge to $(\frac{1}{2})^{|M|}$ from above or from below, if $i \to \infty$. In the next section, we shall study such functions for some particular linear codes.

Let $\mathscr{D}'(f) = \mathscr{D}(f)/\sigma(1)$. Notice that $\mathscr{D}'(f)$ is a purely combinatorial characteristic of f which does not depend on α (we only need $\deg(\alpha) = 1$). Moreover, it is easy to see that $\mathscr{D}'(f) \leqslant |\mathscr{A}|$ and that equality holds here if and only if $f \in \mathscr{A}^{\perp}$. The following theorem gives a concrete example of this.

Let n be a natural number and $M = \mathscr{B}_n$. Hence, $\mathscr{B}_M$ are the ordinary boolean functions of n variables. A Reed–Müller code of order r and length 2^n, denoted as $\mathscr{R}(r, n)$, is the set of all boolean functions, which are expressible as polynomials of order r and n variables over the two element field $\{0, 1\}$ (see [5]).

Theorem 5.8. *If $M = \mathscr{B}_n$ and G is a Reed–Müller code $\mathscr{R}(r, n)$, then*

$$\mathscr{D}'(f) \leqslant \frac{2^n}{2^{r+1}} \prod_{j=0}^{r} \frac{2^n - 2^j}{2^{r+1} - 2^j}$$

and equality holds if and only if $f \in G$.

Proof. $\mathscr{A}$ is the set of the minimum weight functions in the dual code $G^{\perp}$. In this case, $G^{\perp} = \mathscr{R}(n - r - 1, n)$. By the well-known results on the minimum weight functions in a Reed–Müller code, see [5] for example, we obtain an expression of the size of $\mathscr{A}$. This is the upper bound on $\mathscr{D}'(f)$. Moreover, in the case of Reed–Müller codes, $\mathscr{A}$ generates $G^{\perp}$ and hence $\mathscr{A}^{\perp} = (G^{\perp})^{\perp} = G$. This implies the condition of equality, since $\mathscr{D}'(f) = |\mathscr{A}|$ if and only if $f \in \mathscr{A}^{\perp}$. □

6. The linear functions as the starting set

Let n be a natural number and $M = \mathscr{B}_n$. Let G be the set of all the linear boolean functions of n variables, i.e. the functions $c_0 \oplus c_1 x_1 \oplus \cdots \oplus c_n x_n$, where $c_j \in \{0, 1\}$ for $j = 0, \ldots, n$. Hence, G is $\mathscr{R}(1, n)$, the first-order Reed–Müller code (see [5]). In this particular case we shall compute the quantity $\mathscr{D}'(f)$ (defined in the previous section) upon which the asymptotic behaviour of the probability $p_i(f)$ of a given function f depends. We shall express $\mathscr{D}'(f)$ through the Fourier transform of f and using this we shall study the asymptotic behaviour of $p_i(f)$ for particular f.

The symbols x, y, z are used to denote the elements of $\mathscr{B}_n$. By subscripted variables we shall denote the coordinates of them, and by upper indices we shall distinguish different elements. Let $w^{-1}(1) = \{x \in \mathscr{B}_n \mid w(x) = 1\}$.

Let x and y be nonzero and different elements of $\mathscr{B}_n$. Then for every $z \in \mathscr{B}_n$ and every function $f \in G$ the following holds:

$$f(z) \oplus f(z \oplus x) \oplus f(z \oplus y) \oplus f(z \oplus x \oplus y) = 0. \tag{9}$$

The points z, $z \oplus x$, $z \oplus y$, $z \oplus x \oplus y$ of $\mathscr{B}_n$ are pairwise distinct. Let w be a function, which is equal to 1 on each of these four points and to 0 elsewhere. By (9), $\langle w, \tilde{g}_0 \rangle$ is zero with probability 1 and hence $\Delta_0(w) = 1$. It follows that $\operatorname{rank}(G) \leqslant 4$. Clearly, if $f \in G$ then $f \oplus 1 \in G$. If $|w|$ is odd, then $\langle f, w \rangle \neq \langle f \oplus 1, w \rangle$. Hence, if $|w|$ is odd, then $|\Delta_0(w)| < 1$ and, by Theorem 5.6, $\Delta_0(w) = 0$. It follows that $\operatorname{rank}(G)$ is even. Since the constant functions and all the projections x_j are in G, the set G is nondegenerate. Hence, if $|w| = 2$, then $|\Delta_0(w)| < 1$ and, by Theorem 5.6, $\Delta_0(w) = 0$. Getting these facts together, we obtain that $\operatorname{rank}(G) = 4$.

Now, we shall characterize the elements of $\mathscr{A}$, i.e., the functions of the minimum weight in $G^{\perp}$, in a form suitable for our consideration.

Lemma 6.1. *A function w is a minimum weight function in $G^{\perp}$ if and only if $w^{-1}(1) = \{z^1, z^2, z^3, z^4\}$, where z^j are pairwise distinct and $z^1 \oplus \cdots \oplus z^4 = 0_n$.*

Proof. In this proof, the symbol x_j, which denotes the jth coordinate of x, is used also to denote the corresponding projection function defined on $\mathscr{B}_n$. The projections x_j together with the constant function 1 form a basis of G over the two element field $\{0, 1\}$. Hence, for every w we have $w \in G^{\perp}$ if and only if $|w|$ is even and $\langle w, x_j \rangle = 0$ for all $j = 1, \ldots, n$. Clearly, $\langle w, x_j \rangle = \bigoplus_{w(x)=1} x_j$. Hence, $w \in G^{\perp}$ if and only if $|w|$ is even and

$$\bigoplus_{w(x)=1} x = 0_n.$$

Since the minimum weight of $G^{\perp}$ is 4, the statement is proved. □

For the computation of $\mathscr{D}'(f)$ we shall use the discrete Fourier transform of the boolean function f. We shall use the notation from [5].

Definition 6.2. *Let f be a boolean function. Then for every $x, y \in \mathscr{B}_n$, let $F(x) = (-1)^{f(x)}$ and*

$$\hat{F}(y) = \sum_{z \in \mathscr{B}_n} F(z)(-1)^{\langle y, z \rangle}.$$

Notice that $\hat{F}(y) = 2^n S_f(y)$. The particular structure of $\mathscr{A}$ allows to express $\mathscr{D}'(f)$ in the following way.

Lemma 6.3. *Let G be the set of the linear boolean functions. Then for every boolean function f we have*

$$\mathscr{D}'(f) = \frac{1}{3.2^{n+3}} \left(\sum_{y \in \mathscr{B}_n} \hat{F}(y)^4 - 3.2^{3n} + 2^{2n+1} \right).$$

Proof. Notice that if $w^{-1}(1) = \{z^1, \ldots, z^4\}$ and the elements z^j are distinct, then $(-1)^{\langle w, f\rangle} = F(z^1), \ldots, F(z^4)$. By Lemma 6.1, $\mathscr{D}'(f)$ is a sum over all unordered four tuples $\{z^1, \ldots, z^4\}$ satisfying $\bigoplus z^j = 0_n$. We shall express this sum using the sum over all ordered four tuples satisfying the same condition, but the elements of which are not necessarily distinct. Notice that if $\bigoplus z^j = 0_n$ and $z^i = z^j$ for some $i \neq j$, then one of the four following situations takes place: $z^1 = z^2 \neq z^3 = z^4$, $z^1 = z^3 \neq z^2 = z^4$, $z^1 = z^4 \neq z^2 = z^3$, $z^1 = z^2 = z^3 = z^4$. There are $3 \cdot 2^n(2^n - 1) + 2^n$ of such ordered four tuples. For all of them, we have $F(z^1) \ldots F(z^4) = 1$. Every element $w \in \mathscr{A}$ corresponds to 24 ordered four tuples satisfying $\bigoplus z^j = 0_n$. Hence, we obtain

$$\mathscr{D}'(f) = \frac{1}{24}\left(\sum_{\bigoplus z^j = 0_n} F(z^1) \ldots F(z^4) - 3 \cdot 2^{2n} + 2^{n+1} \right). \tag{10}$$

Since

$$\sum_{y \in \mathscr{B}_n} (-1)^{\langle z^1 \oplus \cdots \oplus z^4, y\rangle} = \begin{cases} 2^n & \text{if } z^1 \oplus \cdots \oplus z^4 = 0_n, \\ 0 & \text{otherwise,} \end{cases}$$

we can write the inner sum in (10) in the form

$$\left(\tfrac{1}{2}\right)^n \sum_{y \in \mathscr{B}_n} \sum_{z^1, \ldots, z^4 \in \mathscr{B}_n} F(z^1) \ldots F(z^4)(-1)^{\langle z^1 \oplus \cdots \oplus z^4, y\rangle}$$

$$= \left(\tfrac{1}{2}\right)^n \sum_{y \in \mathscr{B}_n} \left(\sum_{z \in \mathscr{B}_n} F(z)(-1)^{\langle z, y\rangle} \right)^4 = \left(\tfrac{1}{2}\right)^n \sum_{y \in \mathscr{B}_n} \hat{F}(y)^4.$$

This completes the proof. □

The following notion is introduced in [6].

Definition 6.4 (Rothaus [6]). A boolean function f of n variables is called bent if for all $y \in \mathscr{B}_n$ we have $|\hat{F}(y)| = 2^{n/2}$.

Remark. The Hamming distance of an arbitrary boolean function of n variables from the set of all the linear functions of n variables is $2^{n-1} - \max_{y \in \mathscr{B}_n} |\hat{F}(y)|/2$, where $\hat{F}$ is its Fourier transform. Bent functions achieve the maximum of this distance, since every boolean function satisfies $|\hat{F}(y)| \geqslant 2^{n/2}$ for at least one $y \in \mathscr{B}_n$. For the proof and for more details on the bent functions we refer to [6] or [5].

Using bent functions, we can characterize the extremal points of $\mathscr{D}'(f)$ if n is even. For n odd, only the upper bound is tight.

Theorem 6.5. *For every boolean function f of n variables the following inequalities hold:*

$$-\tfrac{1}{12} 2^n(2^n - 1) \leqslant \mathscr{D}'(f) \leqslant \tfrac{1}{24} 2^n(2^n - 1)(2^n - 2).$$

Moreover, the lower bound is attained if and only if n is even and f is bent. The upper bound is attained if and only if f is linear.

Proof. By a straightforward computation using the definition of $\hat{F}(y)$ for $y \in \mathscr{B}_n$, one can verify that

$$\sum_{y \in \mathscr{B}_n} \hat{F}(y)^2 = 2^{2n}.$$

Hence, it follows from the Cauchy inequality that

$$2^{4n} = \left(\sum_{y \in \mathscr{B}_n} 1 \cdot \hat{F}(y)^2 \right)^2 \leqslant 2^n \sum_{y \in \mathscr{B}_n} \hat{F}(y)^4$$

and that equality holds here if and only if $|\hat{F}(y)|$ is equal for all $y \in \mathscr{B}_n$, i.e. if $|\hat{F}(y)| = 2^{n/2}$ for all y. By substitution of the above estimate of the sum of $\hat{F}(y)^4$ into Lemma 6.3, we obtain the lower bound on $\mathscr{D}'(f)$ and, by comparing with the definition of the bent functions, we obtain also the condition of equality. In the case of equality, n is even, because $\hat{F}(y)$ is always a rational number.

The upper bound and the corresponding condition of equality follows from Theorem 5.8 for the order $r = 1$. It may also be proved by the observation that the sum of $\hat{F}(y)^4$ over all $y \in \mathscr{B}_n$ attains its maximum if and only if $|\hat{F}(t)| = 2^n$ for some $t \in \mathscr{B}_n$ and $\hat{F}(y) = 0$ for all $y \neq t$. □

The properties of $\mathscr{D}'(f)$ proved above have the following consequence concerning the behaviour of $p_i(f)$ for large i.

Theorem 6.6. *Let G be the set of the linear boolean functions of n variables and let α be a nonlinear connective of degree* 1. *Then there exists i_0 such that for all $i \geqslant i_0$ we have*

(a) *if n is even, f_1 is bent and f_2 is not bent then $p_i(f_1) < p_i(f_2)$;*

(b) *if f_3 is linear and f_4 is not linear then $p_i(f_3) > p_i(f_4)$.*

Proof. It follows by combining Theorem 6.5 and 5.5. □

Notice also that, since $\sum_f p_i(f) = 1$, Theorem 6.6 implies the following weaker statement: for all i large enough $p_i(f) > (\frac{1}{2})^{2^n}$ for any linear function f and $p_i(f) < (\frac{1}{2})^{2^n}$ for any bent function f, where n is the number of variables of the function f.

7. Open problems

In Theorem 4.8 we were not interested in the values of the constants appearing in the symbol Θ. It would be interesting to find estimates of the corresponding lower and upper bounds as near as possible.

All the results of the present paper are only asymptotic. Namely,in Theorem 4.7 we proved an estimate of $\max_f |p_i(f) - (\frac{1}{2})^{|M|}|$ when i, the number of iterations of our process, tends to infinity. It would be very interesting to obtain an estimate of this for a particular number i of iterations. If i is smaller than $\log_k(|M|/\log_2|G|)$ then there are functions f such that $p_i(f) = 0$, since the number of realizations of $\tilde{g}_i$ is smaller than the number of all the functions $M \to \{0\,1\}$. If i decreases below $\log_k(|M|/\log_2|G|)$ the number of realizations of $\tilde{g}_i$ decreases rapidly and we may expect that for all functions f either $p_i(f) = 0$ or $p_i(f) \gg (\frac{1}{2})^{|M|}$. Hence, in this case, we should estimate $\max_f p_i(f)$ rather than $\max_f |p_i(f) - (\frac{1}{2})^{|M|}|$.

Recently, a partial solution of this problem for particular connectives α was obtained. The estimate obtained in this partial solution implies a close relation between $\sup_i p_i(f)$ for an arbitrary boolean function f and the formula size complexity of f, see [9].

The idea of obtaining estimate of the leading term of $p_i(f) - (\frac{1}{2})^{|M|}$ in Section 5 (Theorem 5.5) takes the advantage of the assumption that α is of degree 1 substantially. Namely, we used this assumption to establish (Lemma 5.2 and 5.3) which of the terms of (2) are the leading terms. If α is of degree at least 2 we are not able to do it at present. Hence, it would be interesting to find methods of finding functions f satisfying $p_i(f) < (\frac{1}{2})^{|M|}$ (or the converse inequality) for all i large enough which apply also to the case of connectives of the degree at least 2.

Acknowledgements

The author is very grateful to professor P. Štěpánek for valuable remarks concerning this paper and support. The author is also very grateful to professor A.A. Sapozhenko from Moscow State University for helpful discussion and also for directing his attention to [5], in particular, to the section related to the bent functions.

The author would also like to thank the referees of Discrete Mathematics for many useful comments and suggestions concerning the paper.

References

[1] C.M. Adams and S.E. Tavares, Generating and counting binary bent sequences, IEEE Trans. Inform. Theory 36 (1990) 1170–1173.

[2] E.R. Berlekamp, Algebraic Coding Theory (McGraw-Hill, New York, 1968).

[3] J. Bruck, Harmonic analysis of polynomial threshold functions, SIAM J. Discrete Math. (1990) 168–177.

[4] J.H. van Lint, Introduction to Coding Theory (Springer, New York, 1982).

[5] F.J. MacWilliams and N.J.A. Sloane, The Theory of Error-Correcting Codes (North-Holland, Amsterdam, 3rd ed., 1977).

[6] O.S. Rothaus, On 'bent' functions, J. Combin. Theory Ser. A 20 (1976) 300–305.

[7] P. Savický, Random Boolean formulas representing every Boolean function with asymptotically equal probability, Discrete Math. 83 (1990) 95–103.
[8] P. Savický, On the bent boolean functions that are symmetric, European J. Combin. 15 (1994) 407–410.
[9] P. Savický, Complexity and Probability of some Boolean Formulas, TR 538, 1994, Univ. Dortmund, Germany.
[10] I. Wegener, The Complexity of Boolean Functions, Wiley-Teubner Series in Computer Science (Teubner, Stuttgart, Wiley, New York, 1987).
[11] R. Yarlagadda and J. Hershey, Analysis and synthesis of bent sequences, IEEE Proc. 136, part E(2) (1989) 112–123.